

Evaluación de la Eficacia de Modelos de Machine Learning para la Toma de Decisiones Basado en Datos Meteorológicos en la Producción Agrícola

Evaluating the Effectiveness of Machine Learning Models for Decision-Making Based on Meteorological Data in Agricultural Production

Nahomi Machuca Catagua¹, Marlon Navia Mendoza², y Luis Cedeño Valarezo³

RESUMEN

Contexto: El clima es uno de los factores que mas suele influir en la producción agrícola. Este estudio evalúa la eficacia de modelos de Machine Learning en la predicción del rendimiento agrícola basado en datos meteorológicos. **Metodología:** Se utilizó una metodología basada en OSEMN (Obtain, Scrub, Explore, Model, Interpret) para el tratamiento de datos y la construcción de modelos, evaluando su desempeño y precisión de la predicción mediante métricas estadísticas. Se analizaron modelos basados en las técnicas Gradient Boosting, Random Forest, XGBoost y SVM. **Resultados:** El modelo XGBoost obtuvo el mejor rendimiento predictivo en términos de precisión (MAE = 0.1104, RMSE = 0.1367), seguido por Gradient Boosting (MAE = 0.1104, RMSE = 0.1370), mientras que SVM presentó la menor capacidad predictiva en este contexto (MAE = 0.1223, RMSE = 0.1529). **Conclusiones:** Estos hallazgos evidencian el potencial de los enfoques basados en árboles de decisión para anticipar cambios en la producción agrícola y mejorar la planificación en función de las condiciones climáticas.

Palabras clave: Predicción agrícola, Modelos predictivos, Clima, Gradient Boosting, Random Forest.

ABSTRACT

Context: Weather is one of the factors that most often influences agricultural production. This study evaluates the effectiveness of Machine Learning models in predicting agricultural yield based on meteorological data. **Methodology:** A methodology based on OSEMN (Obtain, Scrub, Explore, Model, Interpret) was applied for data processing and model construction, evaluating their performance and prediction accuracy through statistical metrics. Models based on Gradient Boosting, Random Forest, XGBoost, and SVM techniques were analyzed. **Results:** The XGBoost model achieved the best predictive performance in terms of accuracy (MAE = 0.1104, RMSE = 0.1367), followed by Gradient Boosting (MAE = 0.1104, RMSE = 0.1370), while SVM showed the lowest predictive capacity in this context (MAE = 0.1223, RMSE = 0.1529). **Conclusions:** These findings highlight the potential of decision tree-based approaches to anticipate production changes and improve planning according to weather conditions.

Keywords: Agricultural prediction, Predictive models, Climate, Gradient Boosting, Random Forest.

Fecha de recepción: Mayo 22, 2025

Fecha de aceptación: Septiembre 30, 2025

Introducción

Hoy en día la producción agrícola moderna ha experimentado una transformación significativa gracias a la incorporación de tecnologías avanzadas para la optimización de recursos y la mejora de la productividad. Sin embargo, la creciente demanda de alimentos, sumada a los desafíos impuestos por el cambio climático, ha impulsado la necesidad de adoptar enfoques innovadores en la agricultura. En este contexto, la integración de técnicas de Machine Learning (ML) y el aprovechamiento de datos meteorológicos se presentan como herramientas fundamentales para optimizar el rendimiento de los cultivos y mejorar la toma de decisiones en la gestión agrícola. Diversos estudios han resaltado la importancia de emplear modelos predictivos para anticipar variaciones en el rendimiento agrícola, siendo los datos meteorológicos uno de los insumos más críticos para lograr predicciones precisas (Van Klompenburg et al., 2020).

La literatura actual evidencia un creciente interés en el uso de modelos de ML para la predicción del rendimiento

de cultivos. Por ejemplo, (Chlingaryan et al., 2018) han mostrado cómo estos enfoques pueden facilitar la estimación del estado del nitrógeno y, en consecuencia, optimizar la aplicación de fertilizantes en sistemas de agricultura de precisión. Asimismo, investigaciones como las de (Oikonomidis et al., 2023) han explorado el potencial del Deep Learning, enfatizando que la incorporación de datos meteorológicos y ambientales mejora significativamente la precisión de los modelos predictivos. Estos estudios, junto con los análisis integrales

¹Facultad de Ciencias de la Vida y Tecnologías, Universidad Laica Eloy Alfaro de Manabí, Manta, Ecuador. E-mail: nahomi.machuca@pg.uleam.edu.ec

²Universidad Técnica de Manabí, Portoviejo, Ecuador. E-mail: marlon.navia@utm.edu.ec

³Escuela Superior Politécnica Agropecuaria de Manabí, Calceta, Ecuador. E-mail: lcedeno@espam.edu.ec



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence.

Cómo citar: Machuca Catagua, N., Navia Mendoza, M., & Cedeño Valarezo, L. (2025). Evaluación de la Eficacia de Modelos de Machine Learning para la Toma de Decisiones Basado en Datos Meteorológicos en la Producción Agrícola. *Ecuadorian Science Journal*, 9(2), 14-20, September–2025. DOI: <https://doi.org/10.46480/esj.9.2.232>

presentados por (Liakos et al., 2018) y (Morellos et al., 2016), han sentado las bases teóricas y metodológicas que sustentan la aplicación de técnicas de ML en la agricultura.

A pesar de estos avances, aún se identifican vacíos científicos importantes, especialmente en lo que respecta a la comparación sistemática de modelos tradicionales de Machine Learning ya que la mayoría de los estudios se han centrado en evaluar la precisión de modelos individuales o en la aplicación de técnicas de Deep Learning, dejando sin explorar en profundidad la comparación sistemática de modelos tradicionales de ML como Support Vector Machines (SVM), XGBoost, Gradient Boosting y Random Forest bajo un marco metodológico riguroso que varía según el contexto y los datos disponibles (García-Arteaga et al., 2020).

En este contexto, el presente estudio se propone evaluar la efectividad de diferentes modelos de Machine Learning en la predicción de la producción agrícola utilizando datos meteorológicos, con el objetivo general de contribuir a tomar decisiones que permitan optimizar el rendimiento de los cultivos. La relevancia de esta investigación está dada por la necesidad de implementar soluciones tecnológicas que se adapten a las condiciones climáticas variables y a las demandas de la agricultura de precisión.

El resto del artículo se encuentra dividido de la siguiente manera: En la sección 2 se analizan los trabajos relacionados, después se describen los materiales y métodos aplicados en la investigación, a continuación, se muestran los resultados y la discusión, y por último conclusiones del estudio.

Trabajos Relacionados

En los últimos años, la aplicación de Machine Learning (ML) en la agricultura ha experimentado un crecimiento significativo, con un enfoque particular en la predicción del rendimiento de cultivos y la toma de decisiones basadas en datos meteorológicos. Diversas investigaciones han explorado el uso de técnicas avanzadas de aprendizaje automático para optimizar la productividad y eficiencia en la agricultura de precisión.

Uno de los estudios más relevantes en este campo es el de (Van Klompenburg et al., 2020), quienes realizaron una revisión sistemática sobre el uso de modelos de ML para la predicción del rendimiento de cultivos, resaltando la importancia de los datos meteorológicos y climáticos como insumos críticos en la mejora de la precisión de estos modelos. Por otro lado, (Liakos et al., 2018) realizaron una revisión general de las aplicaciones de ML en la agricultura, abordando desde la predicción de rendimiento hasta la detección de enfermedades y la gestión de recursos.

A nivel de estudios de caso, (Chlingaryan et al., 2018) enfatizaron la aplicación de técnicas de ML en la estimación del estado del nitrógeno y la predicción de rendimientos, lo que ha permitido mejorar la toma de decisiones en la gestión de cultivos de alta precisión. De manera similar, (Crane Droesch, 2018) evaluó el impacto del cambio climático en la predicción del rendimiento de cultivos mediante métodos de ML, concluyendo que las metodologías semiparamétricas combinadas con redes neuronales profundas pueden mejorar la robustez de las predicciones frente a escenarios climáticos adversos. De

igual manera, (Khaki et al., 2021) demostraron la eficacia de modelos de Boosting, tales como XGBoost y Random Forest, en la predicción del rendimiento del trigo, destacando la importancia de los datos climáticos en la mejora del desempeño de los modelos.

En el campo de la aplicación del Deep Learning, (Oikonomidis et al., 2023) llevaron a cabo un análisis exhaustivo sobre el impacto de estas técnicas en la predicción del rendimiento de cultivos, demostrando que la combinación de datos meteorológicos con modelos basados en redes neuronales convolucionales mejora significativamente las estimaciones.

Además, se han desarrollado estudios específicos para la optimización del uso de recursos agrícolas. (Villavicencio Quintero et al., 2024) propusieron un sistema de toma de decisiones para el riego basado en ML, destacando la utilidad de los bosques aleatorios y los modelos de regresión para calcular la evapotranspiración de los cultivos protegidos. Así mismo, (Santos et al., 2022) desarrollaron un modelo en tiempo real para la gestión del agua en plantaciones de lúpulo, utilizando datos de sensores y técnicas de ML para optimizar el consumo de agua.

Por otro lado, estudios como el de (?) han analizado la predicción agroindustrial mediante el uso de ML con fuentes de información pública, demostrando que la combinación de datos climáticos con series temporales puede mejorar la precisión en la predicción de la producción agrícola. En la misma línea, (?) exploraron el impacto del ML en la predicción del rendimiento de cultivos en Medellín, encontrando que los modelos de PCA Random Forest y PCA XGBoost ofrecen altos niveles de precisión en contextos agrícolas periurbanos.

Investigaciones recientes como la de (?) también destacan que la incorporación de la inteligencia artificial y el aprendizaje automático en el sector agropecuario no solo optimiza la productividad, sino que a su vez impulsa prácticas más sostenibles. Gracias al análisis de datos en tiempo real, estas tecnologías permiten tomar decisiones más informadas sobre el riego, la fertilización y el manejo general de los cultivos, lo cual representa una herramienta estratégica frente a los desafíos del cambio climático.

Estos estudios destacan el papel crucial del ML en la optimización de la producción agrícola y la gestión de recursos, validando la relevancia de la presente investigación. La comparación de distintos modelos, así como su aplicación en diversos contextos agrícolas, representa un área de interés creciente que contribuirá al desarrollo de nuevas estrategias para la agricultura de precisión.

Materiales y Métodos

En este estudio, se empleó un conjunto de datos sobre el impacto climático en la agricultura de varios países de distintos continentes, publicado por (Khalid, 2024) disponible en <https://www.kaggle.com/datasets/talhachoudary/global-agriculture-climate-impact-dataset>. Este conjunto de datos ofrece una visión integral de cómo las variables climáticas afectan la agricultura a nivel global. Incluye información detallada sobre la producción y rendimiento de cultivos, así como variables climáticas como temperatura, precipitación y emisiones de gases de

efecto invernadero. Además, proporciona datos sobre la superficie cultivada, el uso de fertilizantes y pesticidas, y prácticas agrícolas en diferentes regiones del mundo.

Los datos utilizados abarcan el período comprendido entre 1990 y 2024, con un total de 10,000 registros y 15 columnas. No se detectaron valores nulos en ninguna de las variables, lo que garantiza la integridad del conjunto de datos. Asimismo, se verificó la ausencia de registros duplicados, asegurando la unicidad de la información. El conjunto de datos (Tabla 1) está compuesto por variables numéricas (temperatura promedio, precipitación total, emisiones de CO₂, rendimiento de cultivos, entre otras) y variables categóricas (tipo de cultivo, región, país y estrategias de adaptación).

Para el estudio, se adaptó la metodología OSEMN, que comprende cinco etapas fundamentales. En primer lugar, en la etapa de Obtención, se descargó el conjunto de datos desde la plataforma Kaggle, asegurando su integridad y disponibilidad para el análisis.

En la fase de Depuración, no se detectaron valores nulos en ninguna de las variables, lo que garantiza la integridad del conjunto de datos, dando paso al pre-procesamiento de los datos (Tabla 2) para garantizar su compatibilidad con los modelos de aprendizaje automático.

Posteriormente, en la etapa de Exploración, se realizaron análisis estadísticos descriptivos y visualizaciones para comprender la distribución de las variables y detectar patrones o anomalías. Se construyó una matriz de correlación para identificar asociaciones lineales entre las variables (Figura 1).

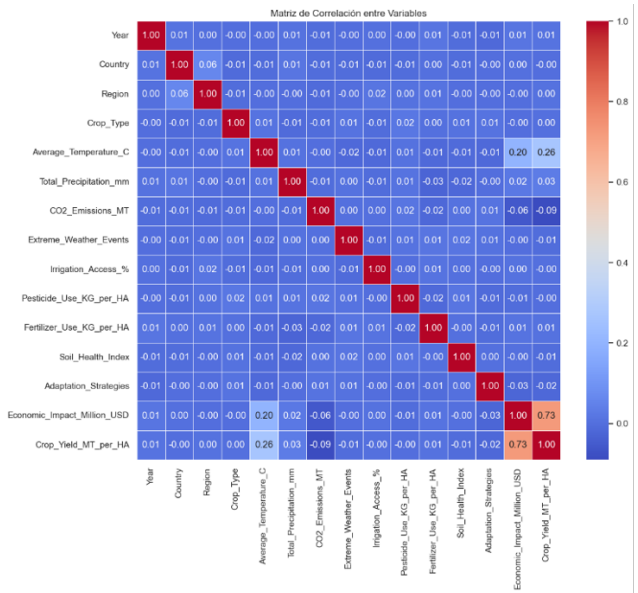


Figura 1. Matriz de correlación

La relación más significativa se observó entre Crop_Yield_MT_per_HA y Economic_Impact_Million_USD, lo que sugiere que el rendimiento agrícola está fuertemente asociado con el impacto económico generado. Otras relaciones destacadas incluyeron una correlación moderada entre Average_Temperature_C y Crop_Yield_MT_per_HA, así como entre Extreme_Weather_Events y la variable objetivo.

En la fase de Modelado todos los modelos fueron entrenados utilizando un conjunto de datos dividido en entrenamiento y prueba, bajo las mismas condiciones, asegurando una comparación justa. Los modelos empleados incluyeron XGBoost, Random Forest Regressor, Gradient Boosting y Support Vector Machines (SVM). Estos modelos fueron seleccionados por su capacidad de aprendizaje sobre relaciones complejas y por su desempeño contrastado en entornos con múltiples variables predictoras, como es el caso del rendimiento agrícola influenciado por datos meteorológicos (Tabla 3).

Para evaluar el desempeño de cada modelo se emplearon métricas como Error Absoluto Medio (MAE), Promedio de los errores al cuadrado (MSE), y Raíz del Error Cuadrático Medio (RMSE). Se descartó el uso del coeficiente de determinación R², dado que no aporta una visión directa sobre el error absoluto y puede inducir a interpretaciones erróneas en conjuntos de datos normalizados (Palma, 2022). Además, se realizaron predicciones a 10 años utilizando datos climáticos proyectados, lo que permitió analizar la capacidad de los modelos para anticipar cambios en el rendimiento agrícola.

Finalmente, en la fase de Interpretación, se analizaron los resultados obtenidos a través de las métricas de desempeño extrayendo conclusiones relevantes sobre la relación entre el clima y la agricultura. Para la implementación del análisis, se utilizó el lenguaje de programación Python con la herramienta Jupyter Notebook, empleando librerías especializadas como Pandas, NumPy, Scikit-learn y Matplotlib para la manipulación, exploración, modelado y visualización de los datos.

Resultados

En esta sección se muestra primero la evaluación de cada modelo para predecir el rendimiento agrícola durante 10 años. Después se presentan los parámetros de desempeño (MAE, MSE y RMSE) de cada modelo evaluado al momento de describir el rendimiento agrícola, con base en los datos de análisis.

Predicción del Rendimiento Agrícola a 10 Años

Para evaluar la capacidad de los modelos en la toma de decisiones agrícolas, se realizaron predicciones a 10 años utilizando datos meteorológicos proyectados. En la Figura 2 se visualizan los valores obtenidos de la predicción de cada modelo.

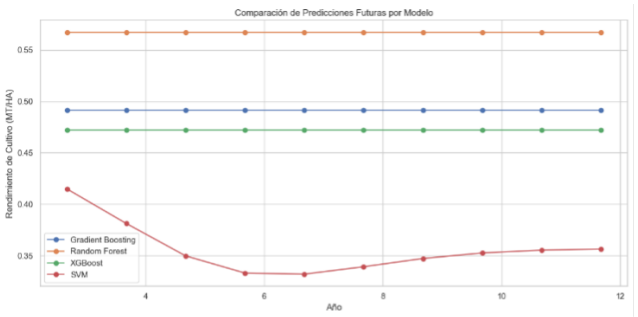


Figura 2. Comparación de Predicciones a 10 años de cada modelo.

Tabla 1. Variables del conjunto de datos utilizado

Variable	Tipo de dato	Descripción
Year	Numérico (entero)	Año de registro de los datos
Country	Texto (categoría)	País al que pertenecen los datos
Region	Texto (categoría)	Región geográfica (ej. Asia, Europa, etc.)
Crop_Type	Texto (categoría)	Tipo de cultivo (trigo, maíz, arroz, etc.)
Average_Temperature_C	Numérico (decimal)	Temperatura media anual en grados Celsius
Total_Precipitation_mm	Numérico (decimal)	Precipitación total anual en milímetros
CO2_Emissions_MT	Numérico (decimal)	Emissiones de CO ₂ en millones de toneladas
Extreme_Weather_Events	Numérico (entero)	Número de eventos climáticos extremos
Irrigation_Access_%	Numérico (decimal)	Porcentaje de tierras con acceso a riego
Pesticide_Use_KG_per_HA	Numérico (decimal)	Uso de pesticidas en kg por hectárea
Fertilizer_Use_KG_per_HA	Numérico (decimal)	Uso de fertilizantes en kg por hectárea
Soil_Health_Index	Numérico (decimal)	Índice de calidad del suelo (escala 0 al 1)
Adaptation_Strategies	Texto (categoría)	Estrategias de adaptación al cambio climático
Economic_Impact_Million_USD	Numérico (decimal)	Impacto económico en millones USD
Crop_Yield_MT_per_HA	Numérico (decimal)	Rendimiento del cultivo en toneladas métricas por hectárea

Fuente: Autores

Tabla 2. Procesamiento de los datos

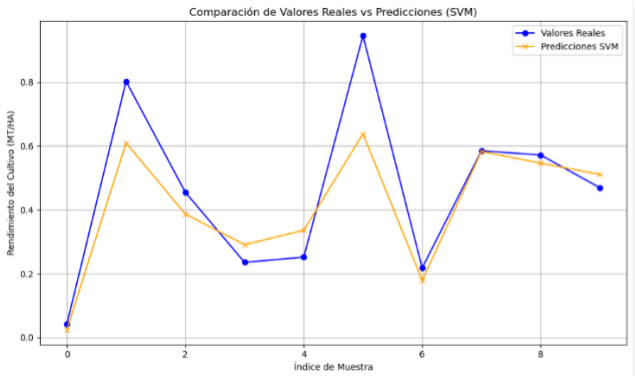
Proceso	Descripción
Conversión de variables categóricas	Transformación a formato numérico mediante One-Hot Encoding
Estandarización de variables numéricas	Aplicación de StandardScaler para homogeneizar la escala
Eliminación de columnas irrelevantes	Se identificaron y eliminaron atributos sin valor predictivo
División de datos	Separación del conjunto en 80% entrenamiento y 20% prueba

Fuente: Autores

Se observa que Random Forest predice sistemáticamente valores más altos (~0.57 MT/HA), lo que sugiere una tendencia a la sobreestimación, alineada con su leve aumento en el error cuadrático. Por otro lado, los modelos Gradient Boosting y XGBoost presentan trayectorias notablemente estables, con valores ligeramente inferiores pero consistentes a lo largo del tiempo, lo que es coherente con sus métricas más bajas de error absoluto y cuadrático. El modelo SVM, en cambio, presenta una curva con mayor variabilidad: inicia con una predicción más alta (~0.41 MT/HA), desciende en los años siguientes hasta ~0.33 y luego comienza una recuperación ligera. Este comportamiento refleja su sensibilidad ante extrapolaciones fuera del rango de entrenamiento, lo cual coincide con sus métricas más elevadas de error.

La predicción de SVM indica un descenso inicial preocupante del rendimiento agrícola, lo cual puede interpretarse como un escenario pesimista o conservador. Refleja una posible pérdida de productividad en el corto plazo, seguida por una

recuperación insuficiente. Sin embargo, se debe considerar que este modelo fue el menos preciso (Figura 3).



Nota. El modelo SVM presenta mayor desviación respecto a los valores reales, en especial en los extremos. Su ajuste es más conservador, lo que explica su rendimiento inferior en las métricas.

Figura 3. Comparación de valores reales vs. predicciones – SVM.

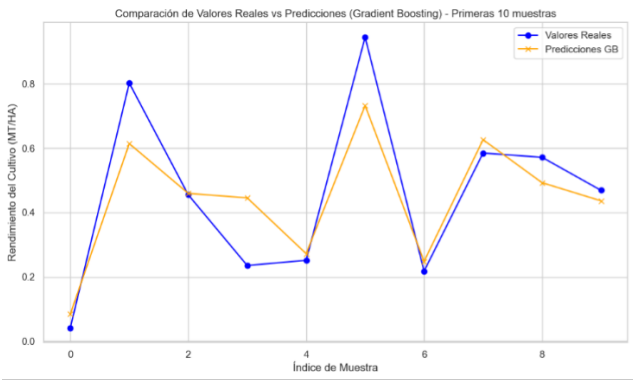
Gradient Boosting muestra un rendimiento moderado alto que se mantiene sin fluctuaciones significativas (Figura 4). Es un escenario positivo, que sugiere una producción estable en los próximos 10 años, dando a entender que el modelo es ideal para planificación agrícola si las condiciones actuales se mantienen.

En la Figura 5 se puede ver la comparación entre la predicción de Random Forest y los datos reales. Si bien este modelo predice el mayor rendimiento agrícola, se debe tener cautela. Dado que el modelo tiende a sobreestimar, este escenario representa un escenario optimista, pero puede no ser tan realista como los de XGBoost y Gradient Boosting.

Tabla 3. Modelos de Machine Learning para la predicción del rendimiento agrícola

Modelo	Descripción
Gradient Boosting	Se ajustó con 1,000 árboles, una tasa de aprendizaje reducida (0.005) y una profundidad máxima de 7
Random Forest	Se configuró con 1000 árboles sin restricción de profundidad y con random_state=42
XGBoost	Fue entrenado con 1000 árboles, tasa de aprendizaje de 0.1, profundidad máxima de 5 y random_state=42
Support Vector Machines (SVM)	Se aplicó con un kernel radial (RBF), requiriendo normalización de los datos

Fuente: Autores



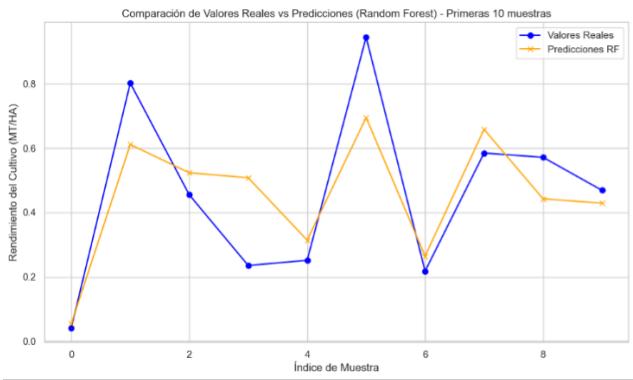
Nota. Predicción estimada constante 0.49 MT/HA, se mantienen muy cercanas a los valores reales, con transiciones suaves y buen seguimiento de picos y valles. Su comportamiento es muy similar al de XGBoost.

Figura 4. Comparación de valores reales vs. predicciones – Gradient Boosting.

La predicción con XGBoost se considera positiva, por lo que se puede confiar en sus estimaciones para tomar decisiones futuras, ya que el margen de error es bajo y consistente (Figura 6). Es importante destacar que, si bien el modelo XGBoost obtuvo las mejores métricas de desempeño, sus predicciones proyectadas para el rendimiento agrícola fueron ligeramente inferiores a las estimadas por el modelo Gradient Boosting. Esta diferencia no contradice su superioridad técnica, sino que responde a la naturaleza conservadora del modelo. Ambos modelos sugieren que, si las condiciones actuales se mantienen, el rendimiento agrícola no sufrirá caídas drásticas, y se ubicará entre 0.47 y 0.49 MT/HA.

Desempeño de los Modelos

La Tabla 4 muestra los parámetros de evaluación de desempeño de los modelos considerados. Como se puede apreciar en esta tabla, aunque los modelos Gradient Boosting y XGBoost obtuvieron un valor idéntico en la métrica MAE (0.1104), el modelo XGBoost demostró un rendimiento global superior. Esta diferencia se evidenció en las métricas MSE (0.0188 en Gradient Boosting vs. 0.0187 en XGBoost) y RMSE (0.1370 vs. 0.1367, respectivamente), donde XGBoost alcanzó valores más bajos, por lo tanto, los resultados muestran que XGBoost fue el modelo con mejor rendimiento global, lo cual lo posiciona como el modelo más preciso y robusto para este caso de estudio.



Nota. Aunque Random Forest sigue bien la tendencia general, tiende a sobreestimar ligeramente en las muestras con valores bajos, lo cual coincide con su comportamiento proyectado.

Figura 5. Comparación de valores reales vs. predicciones – Random Forest.

Tabla 4. Métricas de desempeño de los modelos evaluados

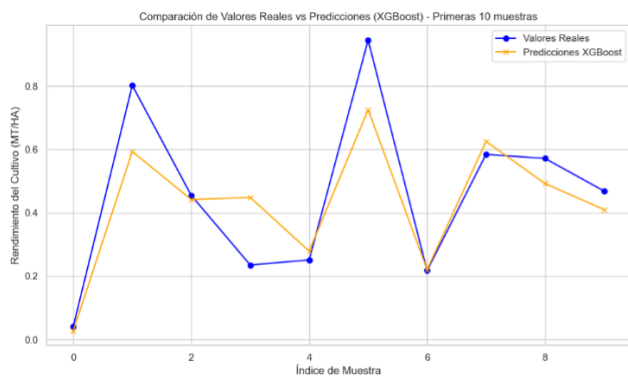
Modelo	MAE	MSE	RMSE
SVM	0.1223	0.0234	0.1529
Gradient Boosting	0.1104	0.0188	0.1370
Random Forest	0.1119	0.0192	0.1385
XGBoost	0.1104	0.0187	0.1367

Fuente: Autores

Discusión

Los hallazgos de este estudio coinciden con investigaciones previas que han demostrado la eficacia de modelos de boosting en la predicción del rendimiento agrícola (Chlingaryan et al., 2018). Se ha observado que XGBoost y Gradient Boosting destacan por su capacidad para manejar datos tabulares con múltiples variables, proporcionando predicciones precisas y estables. Siendo XGBoost el modelo más preciso, con menor error cuadrático medio, lo que refleja una mayor consistencia en sus predicciones. A pesar de que su valor de MAE fue idéntico al de Gradient Boosting, el menor MSE y RMSE indican que cometió errores más pequeños de forma más constante, sin grandes desviaciones. Esto lo posiciona como el modelo más confiable para estimaciones de largo plazo.

El modelo Random Forest, si bien obtuvo un desempeño aceptable, mostró una ligera tendencia a sobreestimar el rendimiento agrícola. Por su parte, el modelo SVM demostró ser menos adecuado para este tipo de



Nota. Las predicciones generadas por XGBoost replican fielmente la tendencia de los valores reales con mínima desviación. Se observan ajustes suaves y consistentes, incluso en los valores extremos. Se puede confiar en sus estimaciones para tomar decisiones futuras, ya que el margen de error es bajo y consistente.

Figura 6. Comparación de valores reales vs. predicciones – XGBoost.

predicciones, lo cual concuerda con estudios previos que han señalado que este modelo puede sobresimplificar relaciones complejas entre múltiples variables climáticas y agrícolas (García-Arteaga et al., 2020).

Las proyecciones futuras a 10 años generadas por los modelos muestran un comportamiento estable del rendimiento agrícola, lo cual puede interpretarse como un escenario favorable siempre que se mantengan las condiciones climáticas y de manejo actuales. XGBoost y Gradient Boosting predicen valores constantes cercanos a 0.47 y 0.49 MT/HA. Es importante aclarar que los valores proyectados, expresados en toneladas métricas por hectárea (MT/HA), representan la cantidad esperada de producción por unidad de superficie cultivada. Por lo que dichos valores equivalen a 470 y 490 kilogramos por hectárea, respectivamente, lo que se considera un rendimiento moderado. Cabe destacar que el análisis se realizó considerando múltiples tipos de cultivos y datos provenientes de distintos países. Esta elección se justificó con el objetivo de capturar una mayor diversidad de condiciones climáticas, prácticas agrícolas y contextos geográficos, lo que permite construir modelos más generalizables y representativos del comportamiento global del sistema agrícola, lo que podría significar una limitación en nuestro estudio, ya que cabe la posibilidad que diversos tipos de cultivos reaccionen de forma diferente al cambio climático o a los eventos naturales. Sin embargo, el enfoque multicultivo y multipaís utilizado en este estudio resulta más apropiado para escenarios de planificación a gran escala y toma de decisiones internacionales.

En conjunto, los resultados respaldan el uso de modelos de aprendizaje automático, particularmente XGBoost, como herramientas efectivas para proyectar el rendimiento agrícola en escenarios complejos. Estas técnicas permiten anticipar tendencias, comparar estrategias y contribuir a la toma de decisiones en contextos de incertidumbre climática, favoreciendo la sostenibilidad y la seguridad alimentaria a largo plazo.

Conclusiones

Los resultados mostraron que XGBoost obtuvo el mejor desempeño, seguido muy de cerca por Gradient Boosting. Ambos mostraron alta precisión y estabilidad en las predicciones. Random Forest, aunque aceptable, tendió a sobreestimar los valores, y SVM fue el menos preciso, con mayor variabilidad.

Las proyecciones futuras a 10 años generadas por los modelos muestran un comportamiento estable del rendimiento agrícola, lo cual puede interpretarse como un escenario favorable siempre que se mantengan las condiciones climáticas y de manejo actuales. XGBoost y Gradient Boosting predicen valores constantes cercanos a 0.47 y 0.49 MT/HA respectivamente, lo que refleja una producción moderadamente alta y sostenida.

Estos hallazgos respaldan la importancia de seleccionar el modelo de Machine Learning adecuado para optimizar la producción agrícola, considerando las condiciones meteorológicas cambiantes. XGBoost es recomendado para escenarios donde la variabilidad climática es un factor determinante, mientras que Gradient Boosting es una opción viable cuando se busca estabilidad en las predicciones.

En conjunto, los resultados respaldan el uso de modelos de aprendizaje automático, particularmente XGBoost, como herramientas efectivas para proyectar el rendimiento agrícola en escenarios complejos. Estas técnicas permiten anticipar tendencias, comparar estrategias y contribuir a la toma de decisiones en contextos de incertidumbre climática, favoreciendo la sostenibilidad y la seguridad alimentaria a largo plazo.

Como trabajo futuro, se podría realizar un análisis diferenciado de los modelos analizados para diversos tipos de cultivo, con el fin de tener un enfoque más preciso para cada producto agrícola.

Agradecimientos

Los autores agradecen a sus respectivas instituciones por el apoyo brindado para la realización de esta investigación.

Referencias

- Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151:61–69.
- Crane Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11).
- García-Arteaga, J. J., Zambrano-Zambrano, J. J., Alcivar-Cevallos, R., and Zambrano-Romero, W. D. (2020). Predicción del rendimiento de cultivos agrícolas usando aprendizaje automático. *Revista Arbitrada Interdisciplinaria Koinonía*, 5(2):144.
- Khaki, S., Pham, H., and Wang, L. (2021). Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports*, 11(1).

- Khalid, T. (2024). Global agriculture climate impact dataset. Kaggle. Data set.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors (Switzerland)*, 18(8).
- Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebensohn, J., Bill, R., and Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-nir spectroscopy. *Biosystems Engineering*, 152:104–116.
- Oikonomidis, A., Catal, C., and Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. *New Zealand Journal of Crop and Horticultural Science*, 51(1):1–26.
- Palma, R. P. (2022). Análisis crítico del coeficiente de determinación (r^2), como indicador de la calidad de modelos lineales y no lineales. *FCNM*, 20(2).
- Santos, R. P., Beko, M., and Leithardt, V. R. Q. (2022). Modelo de machine learning em tempo real para agricultura de precisão. In *Anais da XXII Escola Regional de Alto Desempenho da Região Sul*.
- Van Klompenburg, T., Kassahun, A., and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177.
- Villavicencio Quintero, D., Cabrera Hernández, E., Godo Alonzo, A., and Santana Ching, I. (2024). Sistema para la toma de decisiones en el riego de cultivos protegidos basado en aprendizaje de máquina. *RIELAC*, 44.